

(19)



JAPANESE PATENT OFFICE

## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **05225247 A**(43) Date of publication of application: **03.09.93**

(51) Int. Cl.

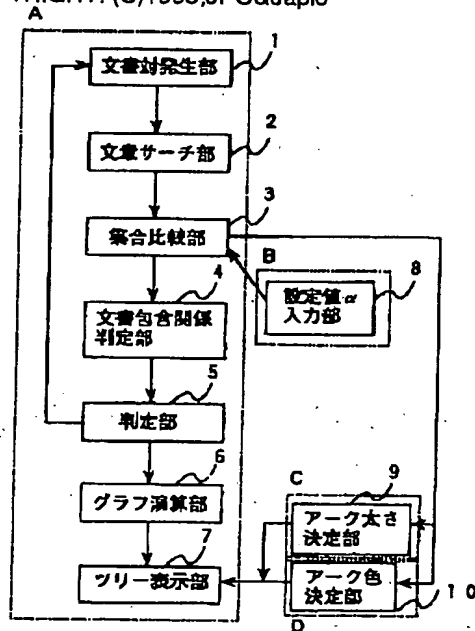
**G06F 15/40****G06F 15/40**(21) Application number: **04004806**(71) Applicant: **NEC CORP**(22) Date of filing: **14.01.92**(72) Inventor: **CHIMURA HIROYASU****(54) INTER-DOCUMENT STRUCTURE DISPLAY METHOD****(57) Abstract:**

**PURPOSE:** To considerably reduce the man-hour by using a keyword to automatically extract relations between plural documents and displaying the structure on a display device.

**CONSTITUTION:** Two documents A and B are selected by a document pair generating part 1, and all sentences in documents A and B are searched by a sentence search part 2 to find keywords registered in a preliminarily prepared set of keywords. As the result, sets  $K_a$  and  $K_b$  of keywords included in documents A and B are settled. A set comparing part 3 compares sets  $K_a$  and  $K_b$  with each other; and when the set  $K_a$  is a subset of the set  $K_b$  or a part of the set  $K_a$  on the outside of the set  $K_b$  is smaller than a set value  $\alpha$ , it is regarded by a document inclusion relation discriminating part 4 that the document B includes the document A, and '1' is substituted into the A-th row and the B-th column of a prepared matrix M. This processing is repeated for all pairs of documents, and the matrix M is regarded as an adjacency matrix in the graph theory by a graph operation part 6, and the known hierarchizing algorithm

is applied to display the structure of the whole of documents as a tree.

COPYRIGHT: (C)1993,JPO&amp;Japlo



(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-225247

(43)公開日 平成5年(1993)9月9日

(51)IntCl.<sup>5</sup>

G 0 6 F 15/40

識別記号

庁内整理番号

FI

技術表示箇所

5 1 0

E

7060-5L

P

7060-5L

5 0 0

C

7060-5L

審査請求 未請求 請求項の数4(全 4 頁)

(21)出願番号 特願平4-4808

(22)出願日 平成4年(1992)1月14日

(71)出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72)発明者 千村 浩靖

東京都港区芝五丁目7番1号日本電気株式  
会社内

(74)代理人 弁理士 京本 直樹 (外2名)

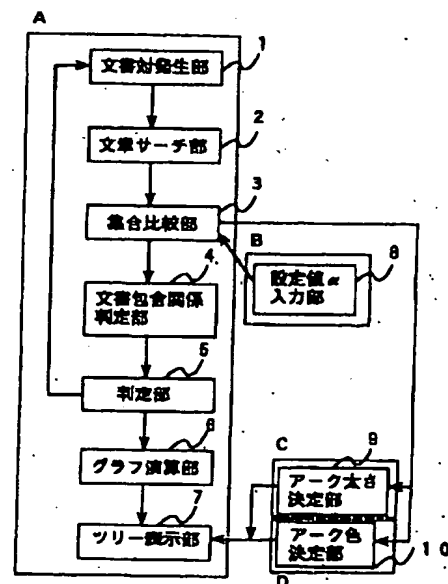
(54)【発明の名称】 文書間構造表示方法

## (57)【要約】

【目的】 複数の文書間の関連を、共通のキーワードの包含状況を利用して自動的に抽出し、構造をディスプレイに表示することにより、文書管理の効率化を図る新しい方法を提案する。

【構成】 1対の文書に対して、あらかじめ用意したキーワード集に登録されているキーワードのうちどれが含まれているかをサーチする文章サーチ部、サーチした結果から文書間の包含関係を判定する文書包含関係判定部、その結果を表示するツリー表示部を中心とする構成をとる。

【効果】 従来、人手で行っていた文書対の関連のあり／なしの判断を自動的に行うことが可能となり、大幅に工数を削減することができるようになった。



1

## 【特許請求の範囲】

【請求項1】 複数の文書間の関連を自動的に抽出し、構造をディスプレイに表示する文書間構造表示方法であって、

任意の2つの文書A、Bに対して、あらかじめ用意したキーワード集に登録されているキーワードの中で、文書Aに含まれるキーワードの集合を $K_a$ 、文書Bに含まれるキーワードの集合を $K_b$ とする第1のステップと、 $K_a$ が $K_b$ の部分集合であるか、またははみ出す部分すなわち集合 $(K_a - K_b)$ の要素数が設定値 $\alpha$ よりも小さい場合には文書Bが文書Aを包含するとみなし、文書Aから文書Bへの方向に関連があると定義する第2のステップと、これをすべての文書対に対して行うことにより、すべての文書間の関連を求める第3のステップと、既知の階層化アルゴリズムを適用して全体の構造をツリーとしてディスプレイに表示する第4のステップとから成ることを特徴とする文書間構造表示方法。

【請求項2】 設定値 $\alpha$ を変化させる第5のステップを有することを特徴とする請求項1記載の文書間構造表示方法。

【請求項3】 ツリーをディスプレイに表示する時、文書AとBの間のアークの太さを、集合 $(K_a - K_b)$ の要素数に応じて変化させることにより、包含関係の強さを表現する第5のステップを有することを特徴とする請求項1記載の文書間構造表示方法。

【請求項4】 ツリーをディスプレイに表示する時、文書AとBの間のアークの色を、集合 $(K_a - K_b)$ の要素数に応じて変化させることにより、包含関係の強さを表現する第5のステップを有することを特徴とする請求項1記載の文書間構造表示方法。

## 【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、複数の文書間の関連を自動的に抽出し、構造をディスプレイに表示することにより文書管理を行うための文書間構造表示方法に関するものである。

【0002】

【従来の技術】 従来、複数の文書間の関連構造を表示するためには、個々の文書を人間が読んで文書対の関連のあり・なしを判定することが必要であった。

【0003】

【発明が解決しようとする課題】 しかしながら、文書が大量になった場合、また、個々の文書のボリュームが大規模である場合、それらすべてを人間が読んで関係を見出すことは大変な工数を必要とし、事実上不可能に近かった。また、例えば文書間の関係づけができたとしても、関係の強弱をも表現できるような表示方法がなかった。

【0004】

【課題を解決するための手段】 本発明は、上記問題点を解決したものであり、第1の発明の文書間構造表示方法

(2)

特開平 5-225247

2

は、複数の文書間の関連を自動的に抽出し、構造をディスプレイに表示する文書間構造表示方法であって、任意の2つの文書A、Bに対して、あらかじめ用意したキーワード集に登録されているキーワードの中で、文書Aに含まれるキーワードの集合を $K_a$ 、文書Bに含まれるキーワードの集合を $K_b$ とする第1のステップと、 $K_a$ が $K_b$ の部分集合であるか、またははみ出す部分すなわち集合 $(K_a - K_b)$ の要素数が設定値 $\alpha$ よりも小さい場合には文書Bが文書Aを包含するとみなし、文書Aから文書Bへの方向に関連があると定義する第2のステップと、これをすべての文書対に対して行うことにより、すべての文書間の関連を求める第3のステップと、既知の階層化アルゴリズムを適用して全体の構造をツリーとしてディスプレイに表示する第4のステップとから成ることを特徴とする。

【0005】 第2の発明の文書間構造表示方法は、第1の発明の文書間構造表示方法において、設定値 $\alpha$ を変化させる第5のステップを有することを特徴とする。

【0006】 第3の発明の文書間構造表示方法は、第1の発明の文書間構造表示方法において、ツリーをディスプレイに表示する時、文書AとBの間のアークの太さを、集合 $(K_a - K_b)$ の要素数に応じて変化させることにより、包含関係の強さを表現する第5のステップを有することを特徴とする。

【0007】 第4の発明の文書間構造表示方法は、第1の発明の文書間構造表示方法において、ツリーをディスプレイに表示する時、文書AとBの間のアークの色を、集合 $(K_a - K_b)$ の要素数に応じて変化させることにより、包含関係の強さを表現する第5のステップを有することを特徴とする。

【0008】

【作用】 本発明は4個の発明からなる。このうち第2、第3、第4の発明は、第1の発明を基本として、それぞれ別個の効果をもたらす新しい機能を付加した発明である。

【0009】 第1の発明は、以下の4つのステップにより、要素間関連構造をディスプレイに表示する。

【0010】 第1のステップ

任意の2つの文書A、B中のすべての文章をサーチし、あらかじめ用意したキーワード集に登録されているキーワードを捜す。今、文書Aに含まれるキーワードの集合が $K_a$ 、文書Bに含まれるキーワードの集合が $K_b$ であったとする。

【0011】 第2のステップ

集合 $K_a$ と $K_b$ の要素を比較する。 $K_a$ が $K_b$ の部分集合であるか、またははみ出す部分すなわち集合 $(K_a - K_b)$ の要素数が設定値 $\alpha$ よりも小さい場合には文書Bが文書Aを包含するとみなし、文書Aから文書Bへの方向に関連があると定義し、用意しておいたマトリックスMの第A行、第B列に“1”を代入する。

50

3

## 【0012】第3のステップ

すべての文書対に対して上記第1～第2のステップを繰り返す。これにより、すべての文書間の関連が求まる。

## 【0013】第4のステップ

第3のステップの結果得られたマトリックスMを、グラフ理論における隣接行列とみなし、既知の階層化アルゴリズムを適用して全体の構造をツリーとしてディスプレイに表示する。

【0014】第2の発明は、第1の発明における設定値 $\alpha$ を変化させることにより、文書間の包含関係のあり／なしの判定の感度を調整する。

【0015】第3の発明は、第1の発明において表示したツリーにおいて、例えば文書A、Bに対する $K_a$ が $K_b$ の部分集合である場合には、文書AからBへのアークの太さを最大にし、一方、 $K_a$ が $K_b$ の部分集合でなく、はみ出す部分がある場合すなわち集合 $(K_a - K_b)$ が空集合でない場合には $(K_a - K_b)$ の要素数が増加するに従ってアークが次第に細くなるように表示する。これにより、文書間のアークの太さにより、文書間の包含関係の強さが一目でわかるようにする。

【0016】第4の発明は、第1の発明において表示したツリーにおいて、例えば文書A、Bに対する $K_a$ が $K_b$ の部分集合である場合には、文書AからBへのアークの色を明るい色または強い色にし、一方、 $K_a$ が $K_b$ の部分集合でなく、はみ出す部分がある場合すなわち集合 $(K_a - K_b)$ が空集合でない場合には $(K_a - K_b)$ の要素数が増加するに従ってアークの色が次第に暗くなる、または弱くなるように表示する。これにより、文書間のアークの色により、文書間の包含関係の強さが一目でわかるようにする。

## 【0017】

【実施例】図1におけるAは、第1の発明の一実施例を示すブロック図である。始めに、文書対発生部1において、処理の対象となる2つの文書A、Bを選択する。次に、文章サーチ部2において、文章A、B中のすべての文章をサーチし、あらかじめ用意したキーワード集に登録されているキーワードを捜す。その結果、文書Aに含まれるキーワードの集合 $K_a$ 、文書Bに含まれるキーワードの集合 $K_b$ が確定する。次に、集合比較部3において、 $K_a$ と $K_b$ の比較を行い、 $K_a$ が $K_b$ の部分集合であるか、または、はみ出す部分がある場合すなわち集合 $(K_a - K_b)$ が空集合でない場合には $(K_a - K_b)$ の要素数が設定値 $\alpha$ よりも小さいか、を調べる。この結果により、次の文書包含関係判定部4にて、文書Aと文書Bの包含関係を判定し、もし包含関係があるならば用

(3)

特開平 5-225247

4

意しておいたマトリックスMの対応する部分に“1”を代入する。次に、判定部5において、すべての文書対に対して処理が終了したか否かを判定し、終了してないならば文書対発生部1に戻り、新たな文書対に対して上記の処理を繰り返す。もし、すべての文書対に対して処理が終了しているのならば、グラフ演算部6において、マトリックスMをグラフ理論における隣接行列とみなし、既知の階層化アルゴリズムを適用して階層化を行う。次のツリー表示部7において、実際に全体の構造をツリーとしてディスプレイに表示する。

【0018】図1におけるAとBは、第2の発明の一実施例を示すブロック図である。設定値 $\alpha$ 入力部8において、設定値 $\alpha$ の値を入力し、この値を集合比較部3に送る。この操作により、文書間の包含関係のあり／なしの判定の感度を調整する。

【0019】図1におけるAとCは、第3の発明の一実施例を示すブロック図である。アーク太さ決定部9は、集合比較部3から集合 $(K_a - K_b)$ の要素数情報を受け取り、各アークの太さを決定し、ツリー表示部に太さ情報を与える。

【0020】図1におけるAとDは、第4の発明の一実施例を示すブロック図である。アーク色決定部10は、集合比較部3から集合 $(K_a - K_b)$ の要素数情報を受け取り、各アークの色を決定し、ツリー表示部に色情報を与える。

## 【0021】

【発明の効果】従来、文書対の関連のあり／なしの判断を手で行っていた方式に対して、本発明による方式によれば、文書対の関連のあり／なしの判断を自動的に行うことが可能となり、大幅に工数を削減することができるといふ効果をもたらす。

## 【図面の簡単な説明】

【図1】第1、2、3及び4の発明の一実施例を示すブロック図

## 【符号の説明】

- 1 文書対発生部
- 2 文章サーチ部
- 3 集合比較部
- 4 文書包含関係判定部
- 5 判定部
- 6 グラフ演算部
- 7 ツリー表示部
- 8 設定値 $\alpha$ 入力部
- 9 アーク太さ決定部
- 10 アーク色決定部

(4)

特開平 5-225247

【図1】

